

(12) INTERNATIONAL APPLICATION PUBLISHED UNDER THE PATENT COOPERATION TREATY (PCT)

(19) World Intellectual Property Organization  
International Bureau



(43) International Publication Date  
18 April 2002 (18.04.2002)

PCT

(10) International Publication Number  
**WO 02/32142 A2**

(51) International Patent Classification<sup>7</sup>: **H04N 7/26**

(21) International Application Number: **PCT/EP01/11393**

(22) International Filing Date: **2 October 2001 (02.10.2001)**

(25) Filing Language: **English**

(26) Publication Language: **English**

(30) Priority Data:  
60/239,661 12 October 2000 (12.10.2000) **US**  
09/887,756 21 June 2001 (21.06.2001) **US**

(71) Applicant: **KONINKLIJKE PHILIPS ELECTRONICS N.V.** [NL/NL]; Groenewoudseweg 1, NL-5621 BA Eindhoven (NL).

(72) Inventors: **VAN DER SCHAAR, Mihaela**; Prof. Holstlaan 6, NL-5656 AA Eindhoven (NL). **RADHA, Hayder**; Prof. Holstlaan 6, NL-5656 AA Eindhoven (NL).

(74) Agent: **LANDOUSY, Christian**; Internationaal Octrooibureau B.V., Prof. Holstlaan 6, NL-5656 AA Eindhoven (NL).

(81) Designated States (*national*): **CN, JP, KR.**

(84) Designated States (*regional*): European patent (AT, BE, CH, CY, DE, DK, ES, FI, FR, GB, GR, IE, IT, LU, MC, NL, PT, SE, TR).

**Published:**

- without international search report and to be republished upon receipt of that report
- entirely in electronic form (except for this front page) and available upon request from the International Bureau

*For two-letter codes and other abbreviations, refer to the "Guidance Notes on Codes and Abbreviations" appearing at the beginning of each regular issue of the PCT Gazette.*



**WO 02/32142 A2**

(54) Title: **SINGLE-LOOP MOTION-COMPENSATION FINE GRANULAR SCALABILITY**

(57) Abstract: A video coding technique having motion compensation for bi-directional predicted frames (B-frames) and predicted frames (P-frames). The video coding technique involves a single-loop prediction-based base layer which includes base layer B- and/or P-frames that are generated from "extended" or "enhanced" base layer reference frames during base layer coding.

## Single-loop motion-compensation fine granular scalability

The present invention relates to video coding, and more particularly to a scalable video coding scheme that employs a single-loop motion compensation process for bi-directional predicted frames (B-frames) and predicted frames and bi-directional predicted frames and (P- and B-frames), while using portions of the enhancement layer frames.

5 Scalable enhancement layer video coding has been used for compressing video transmitted over computer networks having a varying bandwidth, such as the Internet. A current enhancement layer video coding scheme employing fine granular scalable coding techniques (adopted by the ISO MPEG-4 standard) is shown in Fig. 1. As can be seen, the video coding scheme 10 includes a prediction-based base layer 11 coded at a bit rate  $R_{BL}$ , and  
10 an FGS enhancement layer 12 coded at  $R_{EL}$ .

The prediction-based base layer 11 includes intraframe coded I frames, interframe coded P frames which are temporally predicted from previous I- or P-frames using motion estimation-compensation, and interframe coded bi-directional B-frames which are temporally predicted from both previous and succeeding frames adjacent the B-frame using  
15 motion estimation-compensation. The use of predictive and/or interpolative coding i.e., motion estimation and corresponding compensation, in the base layer 11 reduces temporal redundancy therein, but only to a limited extent, since only base layer frames are used for prediction.

The enhancement layer 12 includes FGS enhancement layer I-, P-, and B-  
20 frames derived by subtracting their respective reconstructed base layer frames from the respective original frames (this subtraction can also take place in the motion-compensated domain). Consequently, the FGS enhancement layer I-, P- and B-frames in the enhancement layer are not motion-compensated. (The FGS residual is taken from frames at the same time-instance.) The primary reason for this is to provide flexibility which allows truncation of each  
25 FGS enhancement layer frame individually depending on the available bandwidth at transmission time. More specifically, the fine granular scalable coding of the enhancement layer 12 permits an FGS video stream to be transmitted over any network session with an available bandwidth ranging from  $R_{min} = R_{BL}$  to  $R_{max} = R_{BL} + R_{EL}$ . For example, if the available bandwidth between the transmitter and the receiver is  $B=R$ , then the transmitter

sends the base layer frames at the rate  $R_{BL}$  and only a portion of the enhancement layer frames at the rate  $R_{EL} = R - R_{BL}$ . As can be seen from Fig. 1, portions of the FGS enhancement layer frames in the enhancement layer can be selected in a fine granular scalable manner for transmission. Therefore, the total transmitted bit-rate is  $R = R_{BL} + R_{EL}$ .

5 Because of its flexibility in supporting a wide range of transmission bandwidth with a single enhancement layer.

Fig. 2 shows a block-diagram of a conventional FGS encoder for coding the base layer 11 and enhancement layer 12 of the video coding scheme of Fig. 1. As can be seen, the enhancement layer residual of frame  $i$  ( $FGSR(i)$ ) equals  $MCR(i) - MCRQ(i)$ , where  
10  $MCR(i)$  is the motion-compensated residual of frame  $i$ , and  $MCRQ(i)$  is the motion-compensated residual of frame  $i$  after the quantization and the dequantization processes.

Although the current FGS enhancement layer video coding scheme 10 of Fig. 1 is very flexible, it has the disadvantage that its performance in terms of video image quality is relatively low compared with that of a non-scalable coder functioning at the same  
15 transmission bit-rate. The decrease in image quality is not due to the fine granular scalable coding of the enhancement layer 12 but mainly due to the reduced exploitation of the temporal redundancy among the FGS residual frames within the enhancement layer 12. In particular, the FGS enhancement layer frames of the enhancement layer 12 are derived only from the motion-compensated residual of their respective base layer I-, P-, and B-frames, no  
20 FGS enhancement layer frames are used to predict other FGS enhancement layer frames in the enhancement layer 12 or other frames in the base layer 11.

Accordingly, a scalable video coding scheme is needed that employs portions of the enhancement layer frames for motion compensation to improve image quality while preserving most of the flexibility and attractive characteristics typical to the current FGS  
25 video coding scheme.

The present invention is directed to a video coding scheme that employs portions of enhancement layer frames, typically FGS enhancement layer frames, for predicting B- and P- frames. One aspect of the invention involves a method comprising the steps of: encoding an uncoded video to generate extended base layer reference frames, each  
30 of the extended base layer reference frames including a base layer reference frame and at least a portion of an associated enhancement layer reference frame; predicting base layer frames from the extended base layer reference frames; and generating frame residuals from the uncoded video and the base layer frames.

Another aspect of the invention involves a method comprising the steps of:  
decoding the base layer and enhancement layer streams to generate extended base layer  
reference frames, each of the extended base layer reference frames including a base layer  
reference frame and at least a portion of an associated enhancement layer reference frame;  
5 and predicting base layer frames from the extended base layer reference frames.

Still another aspect of the invention involves a memory medium for encoding  
video, comprising: code for encoding an uncoded video to generate extended base layer  
reference frames, each of the extended base layer reference frames including a base layer  
reference frame and at least a portion of an associated enhancement layer reference frame;  
10 code for predicting base layer frames from the extended base layer reference frames; and  
code for generating frame residuals from the uncoded video and the base layer frames.

A further aspect of the invention involves a memory medium for decoding a  
compressed video having a base layer stream and an enhancement layer stream, comprising:  
code for decoding extended base layer reference frames from the base layer and enhancement  
15 layer streams, each of the extended base layer reference frames including a base layer  
reference frame and at least a portion of an associated enhancement layer reference frame;  
and code for predicting base layer frames from the extended base layer reference frames.

Still a further aspect of the invention involves an apparatus for coding video,  
which comprises: means for encoding an uncoded video to generate extended base layer  
20 reference frames, each of the extended base layer reference frames including a base layer  
reference frame and at least a portion of an associated enhancement layer reference frame;  
means for predicting base layer frames from the extended base layer reference frames; and  
means for generating frame residuals from the uncoded video and the base layer frames.

Still another aspect of the invention involves an apparatus for decoding a  
25 compressed video having a base layer stream and an enhancement layer stream, which  
comprises: means for decoding extended base layer reference frames from the base layer and  
enhancement layer streams, each of the extended base layer reference frames including a base  
layer reference frame and at least a portion of an associated enhancement layer reference  
frame; and means for predicting base layer frames from the extended base layer reference  
30 frames.

The advantages, nature, and various additional features of the invention will  
appear more fully upon consideration of the illustrative embodiments now to be described in

detail in connection with accompanying drawings where like reference numerals identify like elements throughout the drawings:

Fig. 1 shows a current enhancement layer video coding scheme;

Fig. 2 shows a block-diagram of a conventional encoder for coding the base  
5 layer and enhancement layer of the video coding scheme of Fig. 1;

Fig. 3A shows an enhancement layer video coding scheme according to a first exemplary embodiment of the present invention;

Fig. 3B shows an enhancement layer video coding scheme according to a second exemplary embodiment of the present invention;

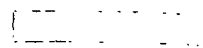
10 Fig. 4 shows a block-diagram of an encoder, according to an exemplary embodiment of the present invention, that may be used for generating the enhancement layer video coding scheme of Fig. 3A;

Fig. 5 shows a block-diagram of an encoder, according to an exemplary embodiment of the present invention, that may be used for generating the enhancement layer  
15 video coding scheme of Fig. 3B;

Fig. 6 shows a block-diagram of a decoder, according to an exemplary embodiment of the present invention, that may be used for decoding the compressed base layer and enhancement layer streams generated by the encoder of Fig. 4;

20 Fig. 7 shows a block-diagram of a decoder, according to an exemplary embodiment of the present invention, that may be used for decoding the compressed base layer and enhancement layer streams generated by the encoder of Fig. 5; and

Fig. 8 shows an exemplary embodiment of a system which may be used for implementing the principles of the present invention.



25

Fig. 3A shows an enhancement layer video coding scheme 30 according to a first exemplary embodiment of the present invention. As can be seen, this enhancement layer video coding scheme 30 includes a prediction-based (single-loop motion compensation) base layer 31 and an enhancement layer 32.

30 The prediction-based base layer 31 includes intraframe coded I-frames and interframe coded P-frames, which are generated conventionally during base layer (non-scalable) coding from standard base layer I- and P- reference frames. In accordance with the principles of the present invention, the prediction-based base layer 31 also includes interframe coded bi-directional B-frames, which are generated from "extended" or

“enhanced” base layer I- and P- or P- and P- reference frames (hereinafter extended base layer I- and P- reference frames) during base layer coding. Each extended base layer reference frame comprises a standard base layer reference frame, and at least a portion of an associated enhancement layer reference frame (one or more bitplanes or fractional bit-planes of the associated enhancement layer reference frame can be used).

The enhancement layer 32 includes I-, P-, and B-frames, which are generated conventionally by subtracting their respective reconstructed (decoded) base layer I-, P- and B- frame residuals from their respective original base layer I-, P- and B-frame residuals. The enhancement layer B-frames are conventionally coded, however, their respective residual frames include less information because their respective base layer frames are partially predicted from enhancement layer reference frames, while the P-frames (and of course the I-frames) are not. The enhancement layer I-, P-, B-frames may be coded with any suitable scalable codec, preferably a fine granular scalable (FGS) codec as shown in Fig. 3A.

As one of ordinary skill in the art will appreciate, the video coding scheme 30 of the present invention improves the image quality of the video. This is because the video coding scheme 30 uses enhancement layer frames to reduce temporal redundancy in the base layer B-frames.

Further, it is important to note that rate-control plays an important role for achieving good performance with the video coding scheme of the present invention. However, even a simplistic approach which allocates the total bit-budget  $B_{tot}$  for a GOP according to  $B_{tot} = b_I * \text{No.}_I\_frames + b_P * \text{No.}_P\_frames + b_B * \text{No.}_B\_frames$ , where  $b_I > b_P > b_B$ , already provides very good results. Further note that a different number of enhancement layer bits/bitplanes (does not have to be an integer number of bits/bitplanes) can be considered for each enhancement layer reference frame used in the motion compensation loop. Moreover, if desired, only certain parts or frequencies within the enhancement layer reference frame need be used as references for the base layer motion-compensation loop.

As should be apparent, the video coding scheme of Fig. 3A modifies the performance of the base layer. The concept of a base layer still exists even though the motion compensation prediction employs enhancement layer information in the motion compensation loop of the base layer, as the resulting B-frame frames are still partially coded with standard quantization and entropy encoding used for conventional base layer encoding. The advantage of coding a portion of the B-frames with standard quantization and entropy encoding is that the resulting video coding scheme remains base layer MPEG-4 compliant.

The remaining portions of the B-frames (B-frame residuals) are coded with a scalable codec such as FGS.

The reason for not coding the entire motion-predicted portion of the enhancement layer I- and P-frames as an "extended" base layer is that the coding scheme employed for the base layer frames is not scalable and if the transmission bandwidth drops below the rate necessary for transmitting this "extended" base layer, the stream cannot be truncated and thus, the entire data until the next I-frame will be lost. With the video coding scheme of the present invention, even though prediction drift occurs if only a very limited part of the enhancement layer bits is received, a decodable stream can still be received at all bit-rates between  $R_{BL}$  and  $R_{max}$ .

Fig. 4 shows a block-diagram of an encoder 40, according to an exemplary embodiment of the present invention, that may be used for generating the enhancement layer video coding scheme of Fig. 3A. As can be seen, the encoder 40 includes a base layer encoder 41 and an enhancement layer encoder 42. The base layer encoder 41 is conventional and includes a motion estimator 43 that generates motion information (motion vectors and prediction modes) from the original video sequence and frames stored in frame memory 60. This motion information is then applied to a motion compensator 44 that generates conventional motion-compensated base layer reference frames and motion-compensated versions of the extended base layer I- and P- reference frames of the present invention (all denoted  $Ref(i)$ ) using the motion information and conventional reference frames and the extended base layer I- and P- reference frames stored in the frame memory 60. A first subtractor 45 subtracts the motion-compensated reference frames  $Ref(i)$  from the original video sequence to generate motion-compensated residuals of the base layer frames  $MCR(i)$ . The motion-compensated residuals of the base layer frames  $MCR(i)$  are processed by a discrete cosine transform (DCT) encoder 46, a quantizer 47, and an entropy encoder 48 to generate the base layer I-, P-, and B-frames, which form a portion of a compressed base layer stream. The motion information generated by the motion estimator 43 is also applied to a multiplexer 49, which combines the motion information with the base layer I-, P-, and B-frames to complete the compressed base layer stream. The quantized motion-compensated residuals of the base layer frames  $MCR(i)$  generated at the output of the quantizer 47 are dequantized by an inverse quantizer 50, and then decoded by an inverse DCT decoder 51. This process generates quantized/dequantized versions of the motion-compensated residuals of the base layer frames  $MCRQ(i)$  at the output of the inverse DCT 51. A first frame flow control device 61 routes the quantized/dequantized motion-compensated residuals of the base

layer I- and P- frames at the output of the inverse DCT 51 to a first adder 52, which sums them with their corresponding motion-compensated base layer reference frames Ref(i), hence generating the conventional base layer reference frames that are stored in the frame memory 60 as described above.

5           The enhancement layer encoder 41 applies the quantized/dequantized motion-compensated residuals of the base layer I-, P-, and B-frames to a second subtractor 53, via a second frame flow control device 59. The second subtractor 53 subtracts the quantized/dequantized motion-compensated residuals of the base layer frames from the corresponding motion-compensated residuals of the base layer frames to generate differential  
10 I-, P-, B-frame residuals FGSR(i). The output of the second subtractor 53 is enhancement layer coded by an FGS encoder 54 or like scalable encoder. The FGS encoder 54 uses conventional DCT encoding followed by conventional bit-plane DCT scanning and conventional entropy encoding to generate a compressed enhancement layer stream. A masking device 55 takes one or more of the coded bit planes of the enhancement layer  
15 encoded I- and P-frames selectively routed through a third frame flow control device 62, and applies this data to a first input 57 of a second adder 56. The quantized/dequantized versions of the motion-compensated residuals of the I- and P-frames MCRQ(i) are routed by the second frame flow control device 59 to a second input 58 of the second adder 56. The second adder 56 generates enhancement layer I- and P- reference frames by summing the one or  
20 more coded bit planes of the enhancement layer encoded I- and P- frames with their respective I- and P- frame residuals MCRQ(i). The enhancement layer I- and P- reference frames outputted by the second adder 56 are directed to the first flow control device 61 which selectively routes them to the first adder 52. The first adder 52 sums the enhancement layer I- and P- reference frames with their corresponding motion-compensated base layer I- and P-  
25 reference frames Ref(i) to generate the extended base layer I- and P- reference frames, which are stored in the frame memory 60 and used as described earlier for predicting the base layer B-frames.

Fig. 3B shows an enhancement layer video coding scheme 100 according to a second exemplary embodiment of the present invention. As can be seen, the video coding  
30 scheme 100 of the second embodiment is substantially identical to the video coding scheme of the first embodiment of Fig. 3A except that the interframe coded predicted P-frames of the base layer 131 are also generated from extended base layer I- or P- reference frames during base layer coding.



The video coding scheme 100 of Fig. 3B provides further improvements in the video image quality. This is because the video coding scheme 100 reduces temporal redundancy in both the P- and the B-frames of the base layer 131 because enhancement layer frames are used in the motion-compensation loop.

5           The video coding schemes of the present invention can be alternated or switched with the current video coding scheme of Fig. 1 for the various portions of a video sequence or for various video sequences. Additionally, switching can be performed among the video coding schemes of Figs. 3A, 3B and the current video coding scheme of Fig. 1, and/or the video coding schemes described in the earlier-mentioned copending U.S. Patent  
10 Applications and/or other video coding schemes. Such switching of video coding schemes can be done based on channel characteristics and can be performed at encoding or at transmission time. Further the video coding schemes of the present invention achieve a large gain in coding efficiency with only a limited increase in complexity.

Fig. 5 shows a block-diagram of an encoder 140, according to an exemplary  
15 embodiment of the present invention, that may be used for generating the enhancement layer video coding scheme of Fig. 3B. As can be seen, the encoder 140 of Fig. 5 is substantially identical to the encoder 40 of Fig. 4 (which is used for generating the enhancement layer video coding scheme of Fig. 3A), except that the frame flow control device 61 used in the encoder 40 is omitted. The frame flow control device 61 is not necessary in this encoder 140  
20 because motion compensation for both the base layer P- frames and the base layer B-frames is based on extended base layer I- and P- reference frames, which contain at least a portion of the enhancement layer coding. Hence, the first adder 52 only generates extended base layer I- and P- reference frames, which are stored in the frame memory 60 (no conventional base layer reference frames are generated or stored in the frame memory 60).

25           Fig. 6 shows a block-diagram of a decoder 70, according to an exemplary embodiment of the present invention, that may be used for decoding the compressed base layer and enhancement layer streams generated by the encoder 40 of Fig. 4. As can be seen, the decoder 70 includes a base layer decoder 71 and an enhancement layer decoder 72. The base layer decoder 71 includes a demultiplexer 73 which receives the encoded base layer  
30 stream and demultiplexes the stream into first and second data streams 74, 75. The first data stream 74, which includes motion information (motion vectors and motion prediction modes) is applied to the motion compensator 76. The motion compensator 76 uses the motion information and the video frames stored in an associated base layer frame memory 77 to reconstruct conventional base layer (I- and P-) reference frames and the extended base layer

reference (I- and P-) frames that are applied to a first input 79 of a first adder 78. The second data stream 75 is applied to a base layer variable length code decoder 81 for decoding, and to an inverse quantizer 82 for dequantizing. The dequantized coefficients are applied to an inverse discrete cosine transform decoder 83 where the dequantized code is transformed into the base layer frame residuals which are applied to a second input 80 of the first adder 78. The first adder 78 sums the base layer P- frame residuals with their respective base layer reference frames and outputs the motion-predicted P- frames. (The base layer I-frame residuals are outputted by the first adder 78 as base layer I-frames.) The first adder 78 also sums the base layer B-frame residuals with their respective extended base layer (I- and P-) reference frames to generate the motion-predicted B-frames. The I- and P- base layer frames outputted by the first adder 78 are stored in the base layer frame memory 77. Additionally, the I-, P-, and B-frames outputted by the first adder 78 may be optionally outputted as a base layer video.

The enhancement layer decoder 72 includes an FGS bit-plane decoder 84 or like scalable decoder that decodes the compressed enhancement layer stream to generate differential I-, P-, and B- frame residuals (enhancement layer frames), which are applied to a second adder 90. The I- and P- frame residuals are also selectively routed by a first frame flow control device 85 to a masking device 86 that takes one or more of the reconstructed enhancement-layer bit-planes (or fractions thereof) of the frame residuals and applies them to a first input 88 of a third adder 87. The base layer (I- and P-) frames outputted by the first adder 78 of the base layer decoder 71 are selectively routed by a second frame flow control device 91 to a second input 89 of the third adder 87. The third adder 87 sums the (I- and P-) enhancement-layer bit-planes with the base layer (I- and P-) frames to generate the extended base layer reference (I- and P-) frames, which are stored in the frame memory 77 and used in the base layer decoder 71 for predicting the base layer B-frames as described above. The second adder 90 sums the enhancement layer I-, P-, and B-frames with their respective base layer I-, P-, and B-frames to generate an enhanced video.

Fig. 7 shows a block-diagram of a decoder 170, according to an exemplary embodiment of the present invention, that may be used for decoding the compressed base layer and enhancement layer streams generated by the encoder 140 of Fig. 5. As can be seen, the decoder 170 of Fig. 7 is substantially identical to the decoder 70 of Fig. 6, except that the base layer frames generated in the first adder 78 are not stored in the base layer frame memory 77. This is not necessary because both the P- and B-frames are predicted using only

the extended base layer reference (I- and P-) frames stored in the frame memory 77. Accordingly, no base layer reference frames are stored in the frame memory 77.

Fig. 8 shows an exemplary embodiment of a system 200 which may be used for implementing the principles of the present invention. The system 200 may represent a television, a set-top box, a desktop, laptop or palmtop computer, a personal digital assistant (PDA), a video/image storage device such as a video cassette recorder (VCR), a digital video recorder (DVR), a TiVO device, etc., as well as portions or combinations of these and other devices. The system 200 includes one or more video/image sources 201, one or more input/output devices 202, a processor 203 and a memory 204. The video/image source(s) 201 may represent, e.g., a television receiver, a VCR or other video/image storage device. The source(s) 201 may alternatively represent one or more network connections for receiving video from a server or servers over, e.g., a global computer communications network such as the Internet, a wide area network, a metropolitan area network, a local area network, a terrestrial broadcast system, a cable network, a satellite network, a wireless network, or a telephone network, as well as portions or combinations of these and other types of networks.

The input/output devices 202, processor 203 and memory 204 may communicate over a communication medium 205. The communication medium 205 may represent, e.g., a bus, a communication network, one or more internal connections of a circuit, circuit card or other device, as well as portions and combinations of these and other communication media. Input video data from the source(s) 201 is processed in accordance with one or more software programs stored in memory 204 and executed by processor 203 in order to generate output video/images supplied to a display device 206.

In a preferred embodiment, the coding and decoding employing the principles of the present invention may be implemented by computer readable code executed by the system. The code may be stored in the memory 204 or read/downloaded from a memory medium such as a CD-ROM or floppy disk. In other embodiments, hardware circuitry may be used in place of, or in combination with, software instructions to implement the invention. For example, the elements shown in Figs. 4-6 may also be implemented as discrete hardware elements.

While the present invention has been described above in terms of specific embodiments, it is to be understood that the invention is not intended to be confined or limited to the embodiments disclosed herein. For example, other transforms besides DCT can be employed, including but not limited to wavelets or matching-pursuits. In another example, other embodiments of the invention may employ only motion compensation in the base layer

for just the P-frames. These and all other such modifications and changes are considered to be within the scope of the appended claims.

11/11/11

11/11/11

## CLAIMS:

1. A method of coding video, comprising the steps of:

encoding (41, 141, 42, 142) an uncoded video to generate extended base layer reference frames, each of the extended base layer reference frames including a base layer reference frame and at least a portion of an associated enhancement layer reference frame;

5 predicting (41, 141) base layer frames from the extended base layer reference frames; and

generating (53) frame residuals from the uncoded video and the base layer frames.

10 2. A method of coding video according to claim 1, further comprising the step of coding (54) the frame residuals with a scalable codec to generate enhancement layer frames.

3. A method of coding video according to claim 1, further comprising the step of coding (54) the frame residuals with a fine granular scalable codec to generate fine granular  
15 scalable enhancement layer frames.

4. A method of coding video according to claim 1, wherein the base layer frames include base layer B-frames.

20 5. A method of coding video according to claim 4, wherein the base layer frames further include P-frames.

6. A method of coding video according to claim 1, wherein the base layer frames include P-frames.

25 7. A method of decoding a compressed video having a base layer stream and an enhancement layer stream, the method comprising the steps of:

decoding (71, 171, 72, 172) the base layer and enhancement layer streams to generate extended base layer reference frames, each of the extended base layer reference

frames including a base layer reference frame and at least a portion of an associated enhancement layer reference frame; and

predicting (71, 171) base layer frames from the extended base layer reference frames.

5

8. A method of decoding video according to claim 7, wherein the decoding (71, 171, 72, 172) step further generates enhancement layer frames; and further comprising the step of generating an enhanced video from the base layer frames and the enhancement layer frames.

10

9. A method of decoding video according to claim 7, wherein the base layer frames include B-frames.

10. A method of decoding video according to claim 9, wherein the base layer frames further include P-frames.

15

11. A method of decoding video according to claim 7, wherein the base layer frames include P-frames.

20 12. A method of decoding video according to claim 7, further comprising the step of generating a video from the base layer frames.

13. A memory medium for encoding video, the memory medium comprising:  
code (41, 141, 42, 142) for encoding an uncoded video to generate extended  
25 base layer reference frames, each of the extended base layer reference frames including a base layer reference frame and at least a portion of an associated enhancement layer reference frame;

code (41, 141) for predicting base layer frames from the extended base layer reference frames; and

30 code (53) for generating frame residuals from the uncoded video and the base layer frames.

14. A memory medium for encoding video according to claim 13, further comprising code (54) for scalable encoding the frame residuals.

15. A memory medium for encoding video according to claim 13, further comprising code (54) for fine granular scalable encoding the frame residuals.

5 16. A memory medium for encoding video according to claim 13, wherein the base layer frames include B-frames.

17. A memory medium for encoding video according to claim 16, wherein the base layer frames further include P-frames.

10

18. A memory medium for encoding video according to claim 13, wherein the base layer frames include P-frames.

19. A memory medium for decoding a compressed video having a base layer  
15 stream and an enhancement layer stream, the memory medium comprising:  
code (71, 171, 72, 172) for decoding extended base layer reference frames  
from the base layer and enhancement layer streams, each of the extended base layer reference  
frames including a base layer reference frame and at least a portion of an associated  
enhancement layer reference frame; and  
20 code for predicting base layer frames from the extended base layer reference  
frames.

20. A memory medium for decoding a compressed video according to claim 19,  
wherein the code (71, 171, 72, 172) for decoding extended base layer reference frames  
25 further generates enhancement layer frames; and further comprising code for generating an  
enhanced video from the base layer frames and the enhancement layer frames.

21. A memory medium for decoding a compressed video according to claim 19,  
wherein the base layer frames include B-frames.

30

22. A memory medium for decoding a compressed video according to claim 21,  
wherein the base layer frames further include P-frames.

23. A memory medium for decoding a compressed video according to claim 19, wherein the base layer frames include P-frames.

24. A memory medium for decoding a compressed video according to claim 19,  
5 further comprising code for generating a video from the base layer frames.

25. An apparatus (40, 140) for coding video, the apparatus comprising:  
means (41, 141, 42, 142) for encoding an uncoded video to generate extended  
base layer reference frames, each of the extended base layer reference frames including a  
10 base layer reference frame and at least a portion of an associated enhancement layer reference  
frame;  
means (41, 141) for predicting base layer frames from the extended base layer  
reference frames; and  
means (53) for generating frame residuals from the uncoded video and the  
15 base layer frames.

26. An apparatus (40, 140) for coding video according to claim 25, further  
comprising means (54) for scalable encoding the frame residuals.

20 27. An apparatus (40, 140) for coding video according to claim 25, further  
comprising means (54) for fine granular scalable encoding the frame residuals.

28. An apparatus (40, 140) for coding video according to claim 25, wherein the  
base layer frames include B-frames.

25 29. An apparatus (40, 140) for coding video according to claim 28, wherein the  
base layer frames further include P-frames.

30. An apparatus (40, 140) for coding video according to claim 25, wherein the  
30 base layer frames include P-frames.

31. An apparatus (70, 170) for decoding a compressed video having a base layer  
stream and an enhancement layer stream, the apparatus comprising:



means (71, 171, 72, 172) for decoding extended base layer reference frames from the base layer and enhancement layer streams, each of the extended base layer reference frames including a base layer reference frame and at least a portion of an associated enhancement layer reference frame; and

5 means (71, 171) for predicting base layer frames from the extended base layer reference frames.

32. An apparatus (70, 170) for decoding a compressed video according to claim 31, wherein the means (71, 171, 72, 172) for decoding extended base layer reference frames  
10 further generates enhancement layer frames; and further comprising means for generating an enhanced video from the base layer frames and the enhancement layer frames.

33. An apparatus (70, 170) for decoding a compressed video according to claim 31, wherein the base layer frames include B-frames.  
15

34. An apparatus (70, 170) for decoding a compressed video according to claim 33, wherein the base layer frames further include P-frames.

35. An apparatus (70, 170) for decoding a compressed video according to claim 31, wherein the base layer frames include P-frames.  
20

1/8

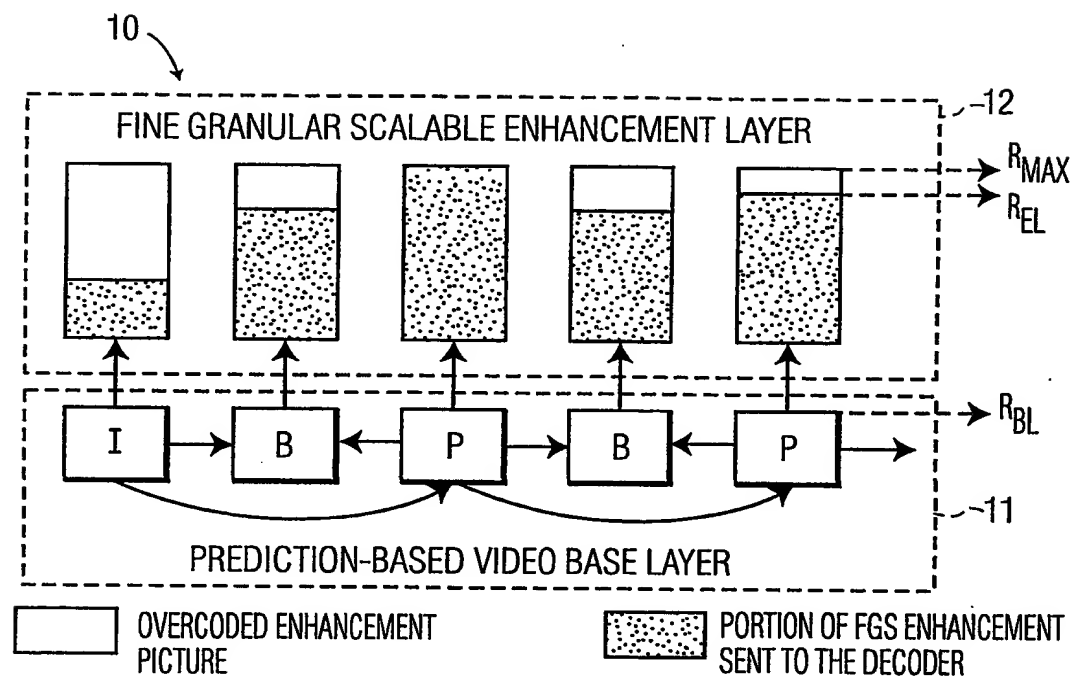


FIG. 1

2/8

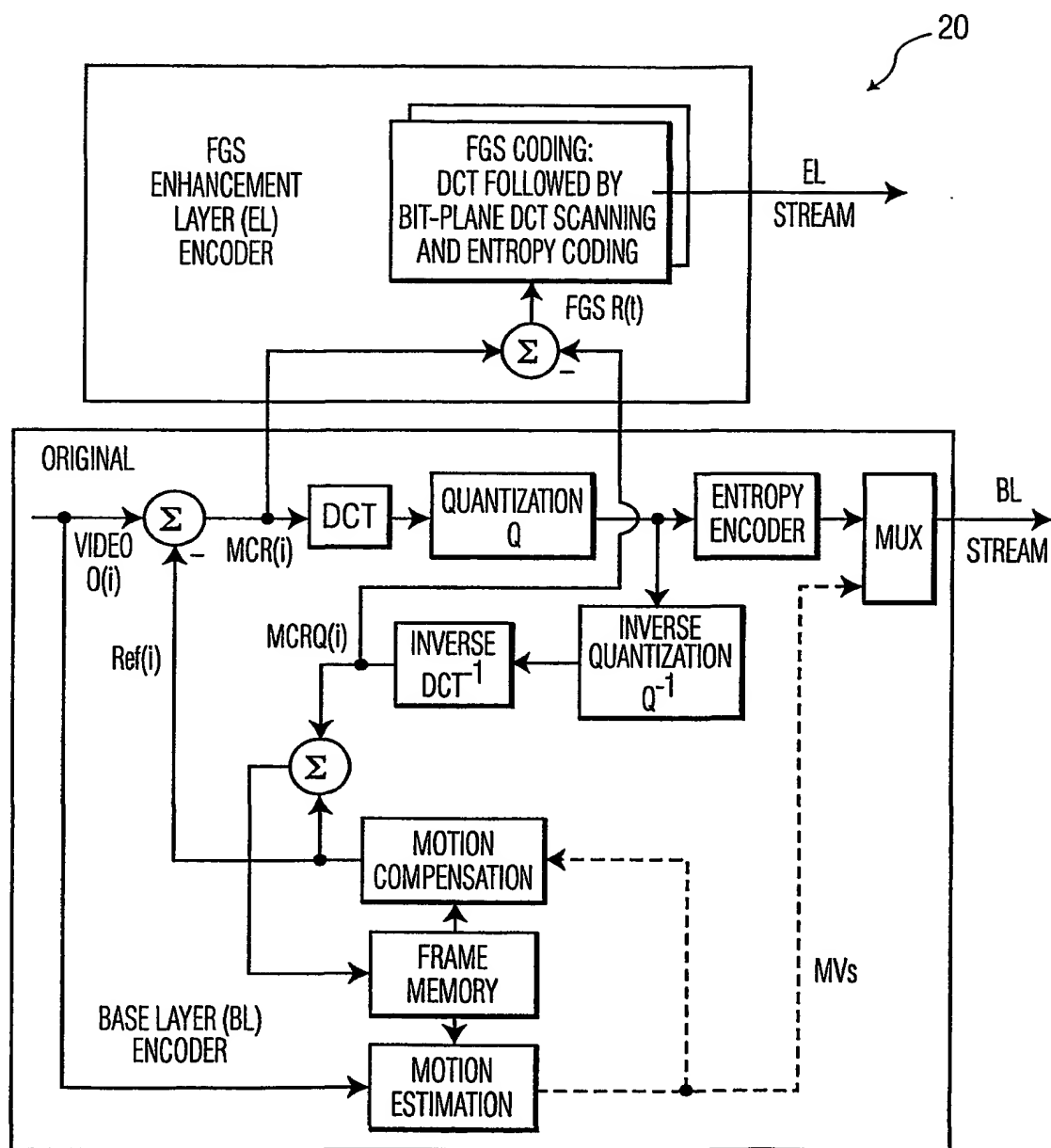


FIG. 2

3/8

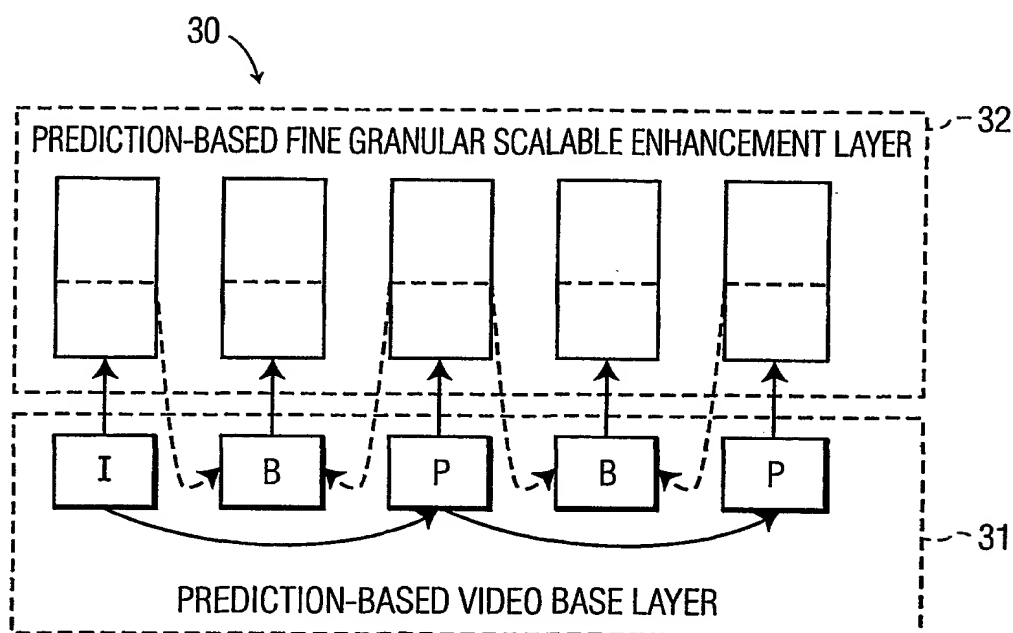


FIG. 3A

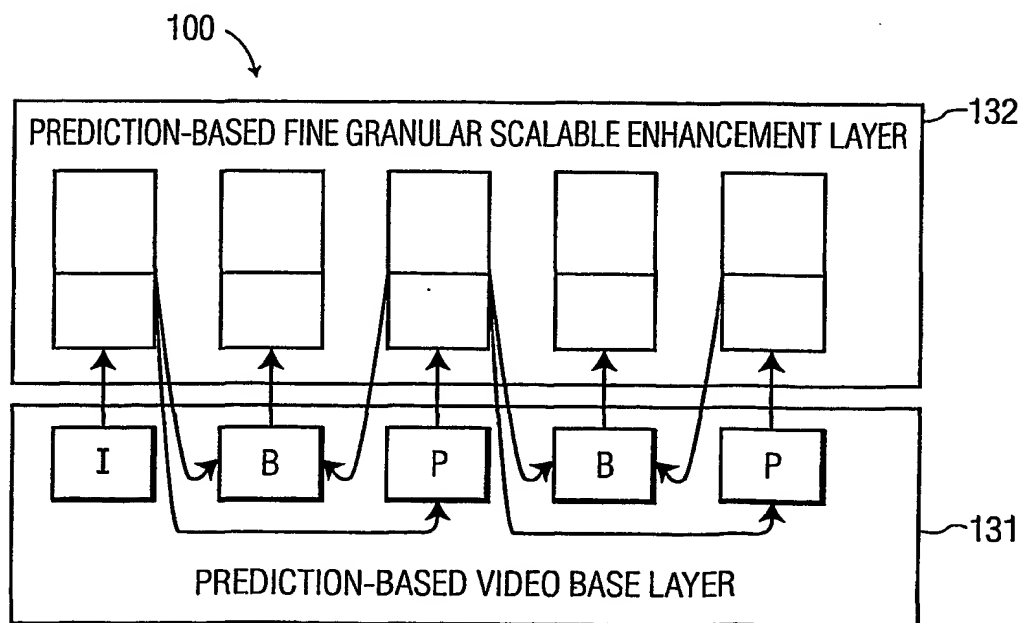


FIG. 3B

4/8

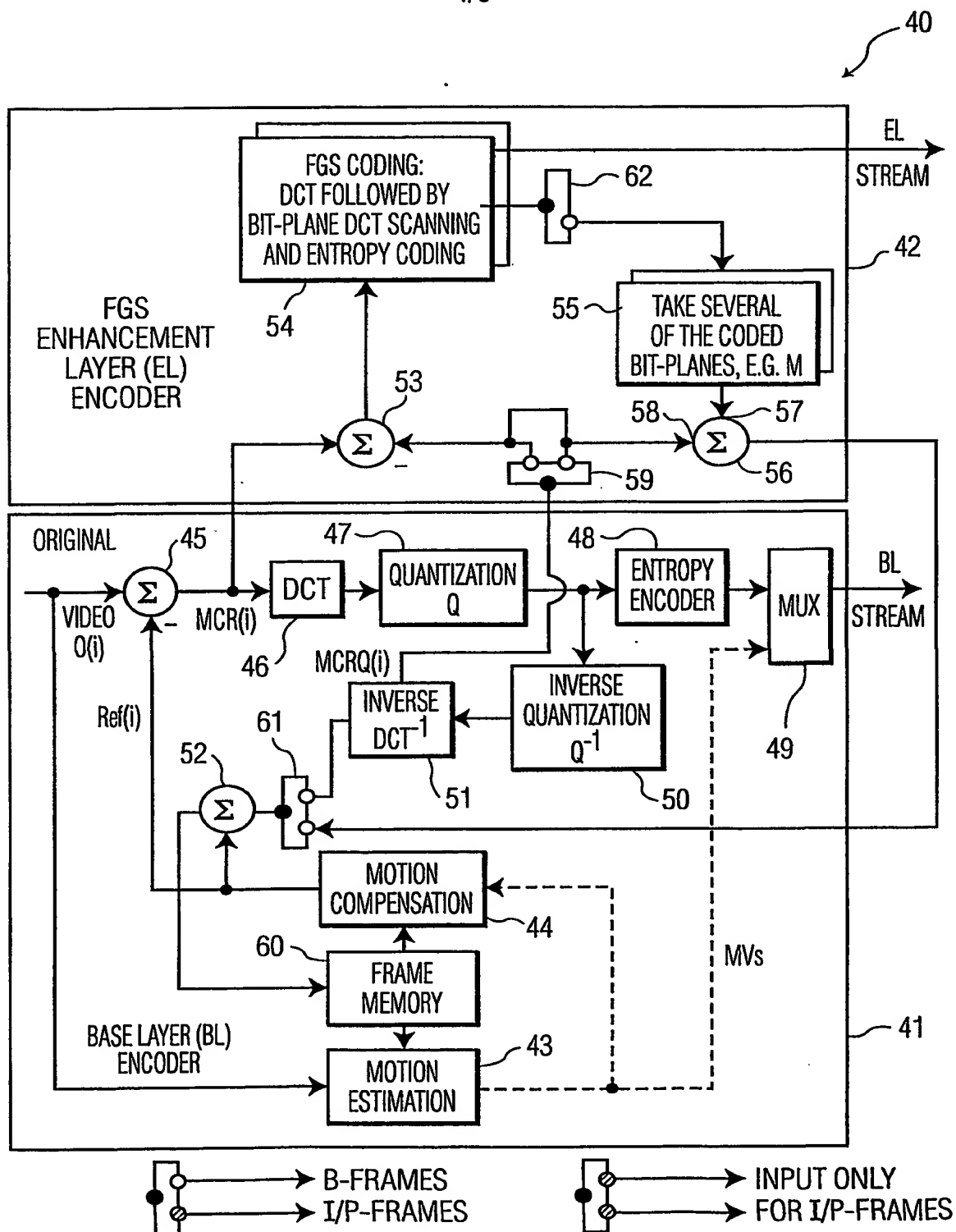


FIG. 4

5/8

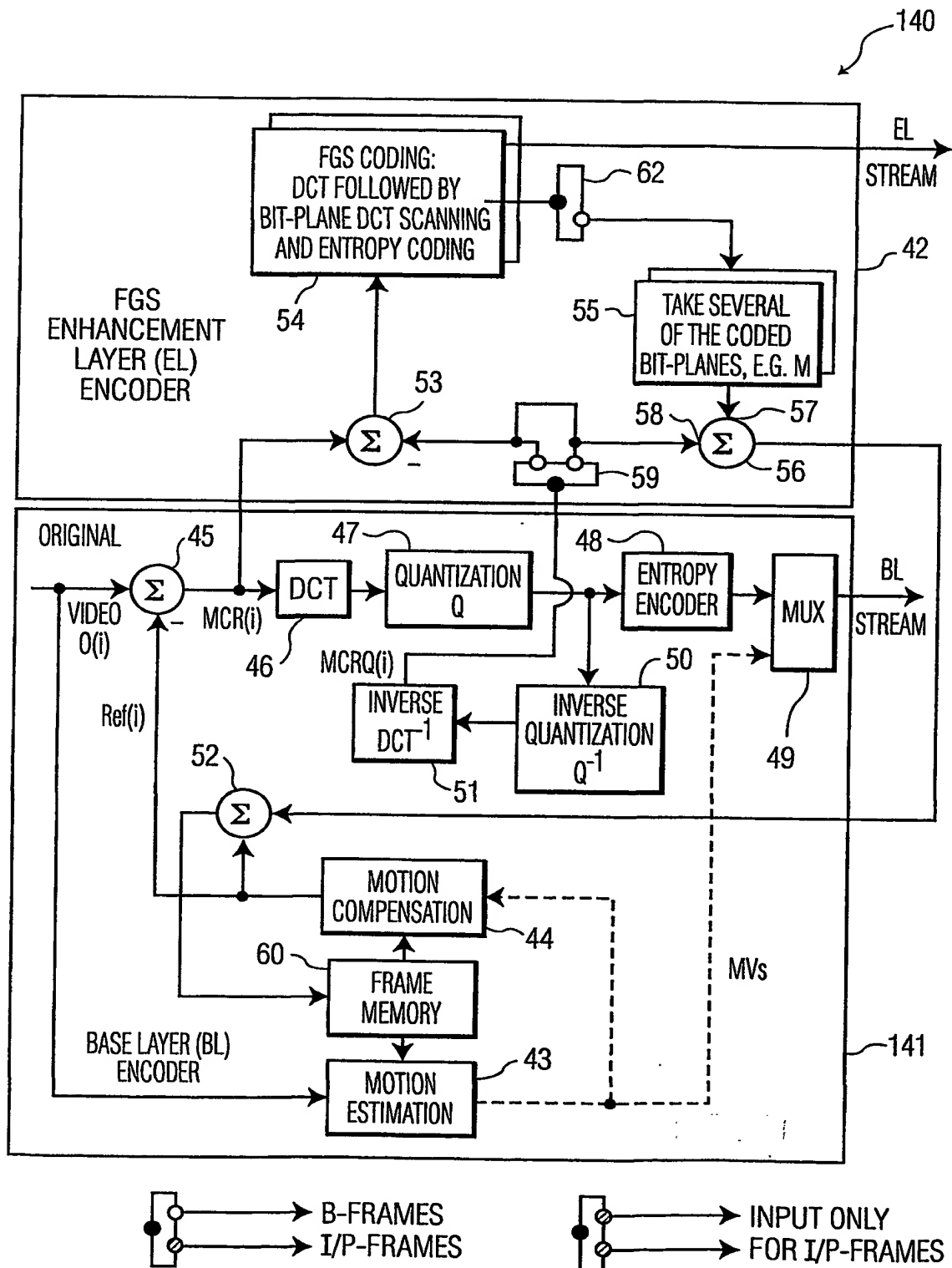


FIG. 5

6/8

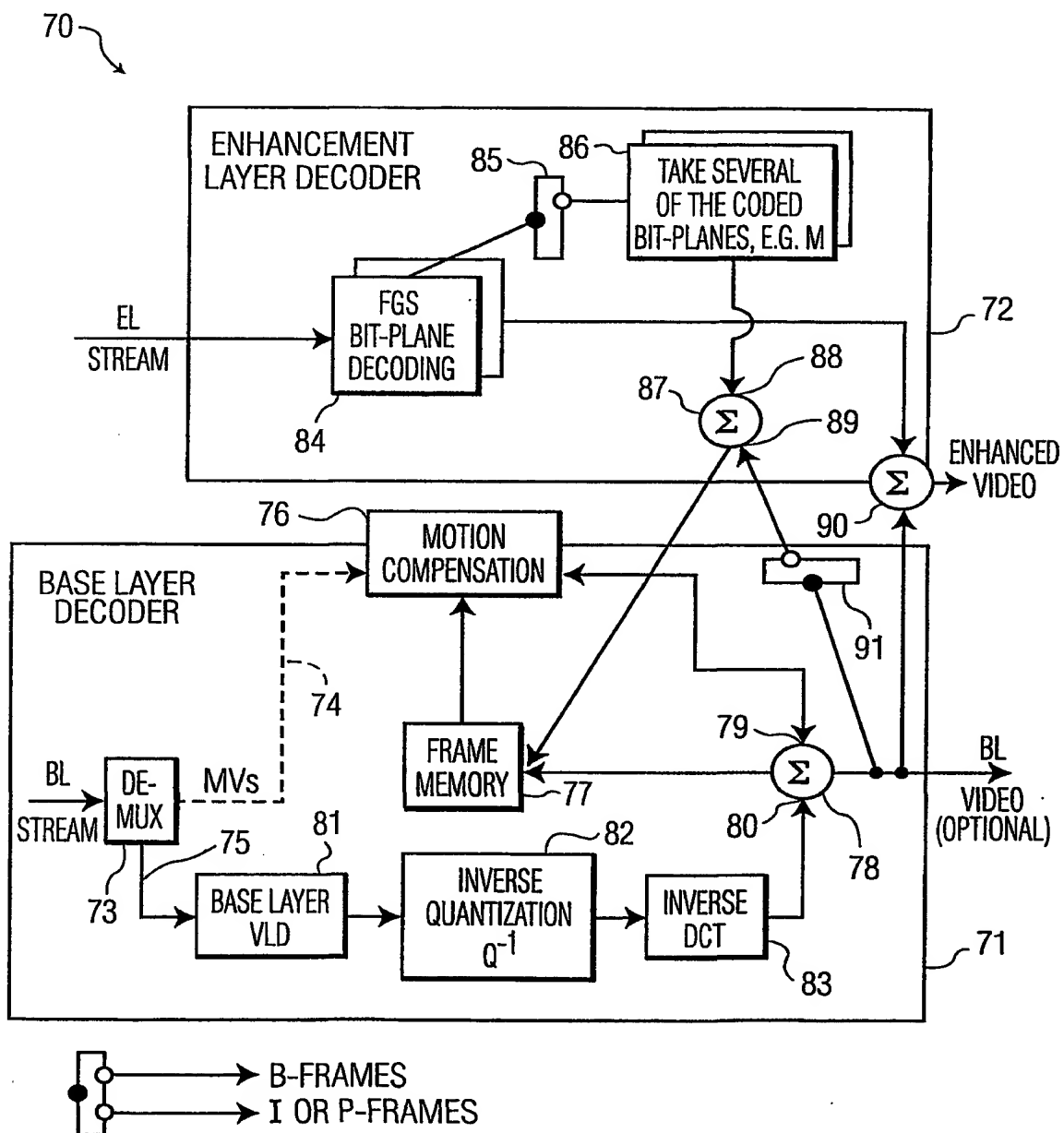


FIG. 6

7/8

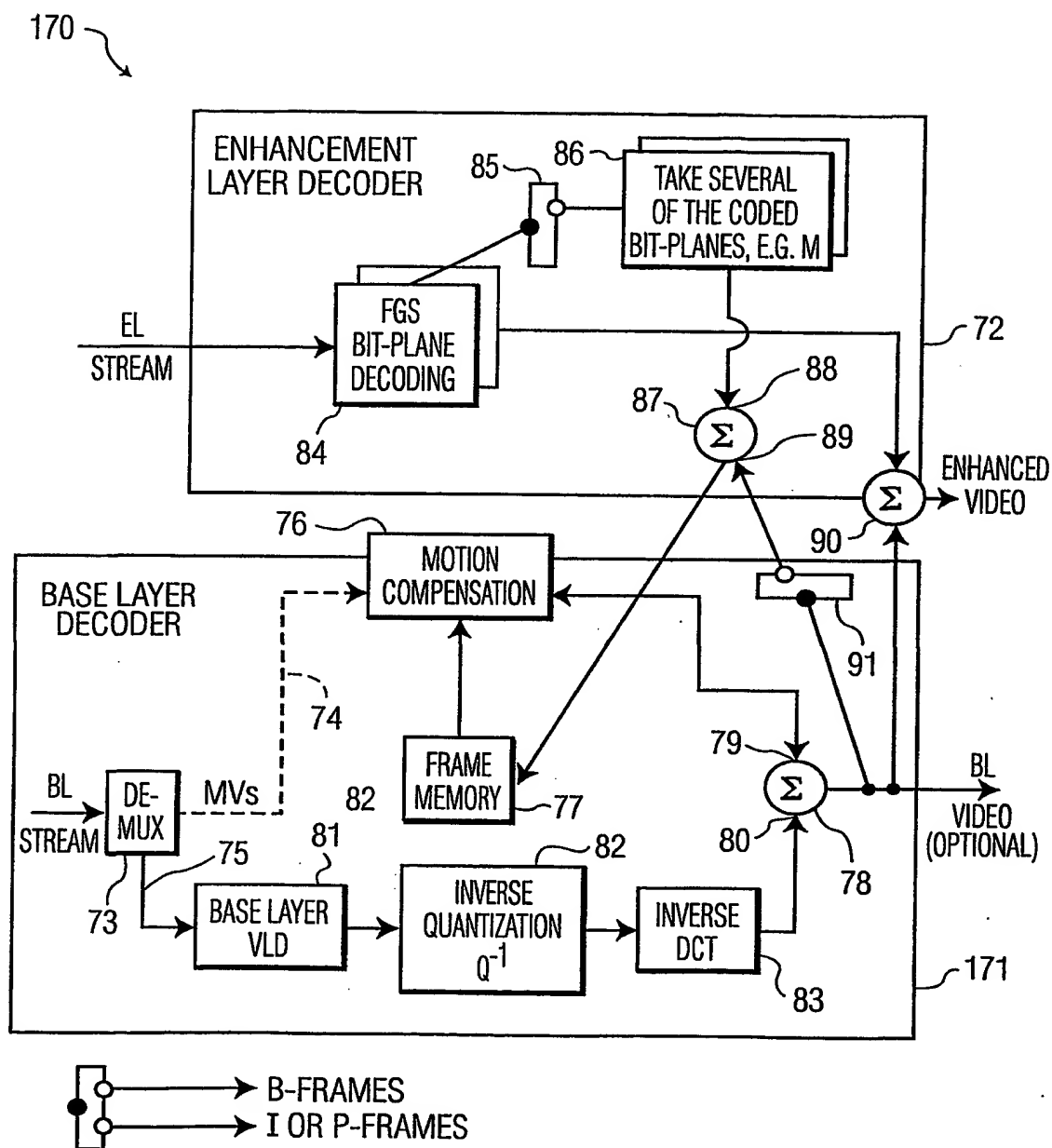


FIG. 7



8/8

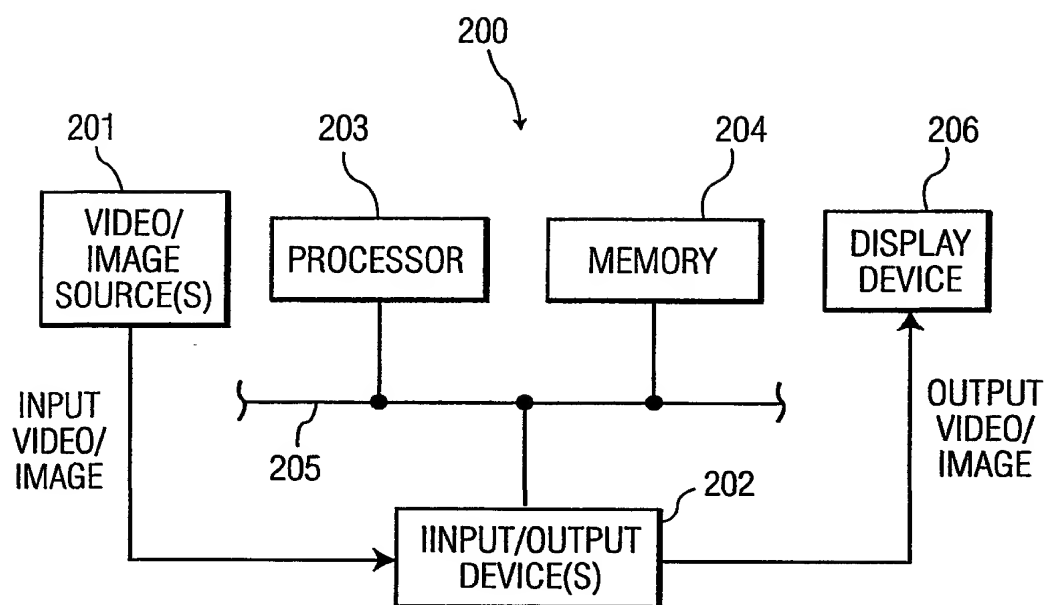


FIG. 8

**THIS PAGE BLANK (USPTO)**

**THIS PAGE**

(12) INTERNATIONAL APPLICATION PUBLISHED UNDER THE PATENT COOPERATION TREATY (PCT)

(19) World Intellectual Property Organization  
International Bureau



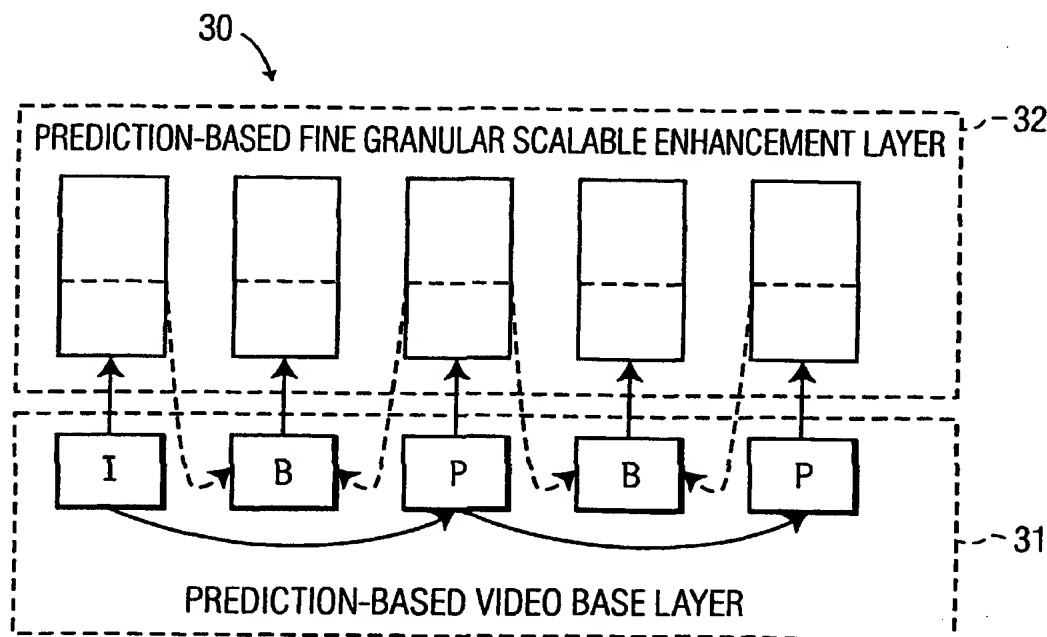
(43) International Publication Date  
18 April 2002 (18.04.2002)

PCT

(10) International Publication Number  
**WO 02/032142 A3**

- (51) International Patent Classification<sup>7</sup>: **H04N 7/50**, 7/26, 7/46
- (21) International Application Number: PCT/EP01/11393
- (22) International Filing Date: 2 October 2001 (02.10.2001)
- (25) Filing Language: English
- (26) Publication Language: English
- (30) Priority Data:  
60/239,661 12 October 2000 (12.10.2000) US  
09/887,756 21 June 2001 (21.06.2001) US
- (71) Applicant: **KONINKLIJKE PHILIPS ELECTRONICS N.V.** [NL/NL]; Groenewoudseweg 1, NL-5621 BA Eindhoven (NL).
- (72) Inventors: **VAN DER SCHAAR, Mihaela**; Prof. Holstlaan 6, NL-5656 AA Eindhoven (NL). **RADHA, Hayder**; Prof. Holstlaan 6, NL-5656 AA Eindhoven (NL).
- (74) Agent: **LANDOUSY, Christian**; Internationaal Octrooibureau B.V., Prof. Holstlaan 6, NL-5656 AA Eindhoven (NL).
- (81) Designated States (*national*): CN, JP, KR.
- (84) Designated States (*regional*): European patent (AT, BE, CH, CY, DE, DK, ES, FI, FR, GB, GR, IE, IT, LU, MC, NL, PT, SE, TR).
- Published:  
— with international search report
- (88) Date of publication of the international search report:  
12 December 2002
- For two-letter codes and other abbreviations, refer to the "Guidance Notes on Codes and Abbreviations" appearing at the beginning of each regular issue of the PCT Gazette.

(54) Title: SINGLE-LOOP MOTION-COMPENSATED FINE GRANULAR SCALABILITY



(57) Abstract: A video coding technique having motion compensation for bi-directional predicted frames (B-frames) and predicted frames (P-frames). The video coding technique involves a single-loop prediction-based base layer which includes base layer B- and/or P-frames that are generated from "extended" or "enhanced" base layer reference frames during base layer coding.

WO 02/032142 A3

## INTERNATIONAL SEARCH REPORT

International Application No

PCT/EP 01/11393

## A. CLASSIFICATION OF SUBJECT MATTER

IPC 7 H04N7/50 H04N7/26 H04N7/46

According to International Patent Classification (IPC) or to both national classification and IPC

## B. FIELDS SEARCHED

Minimum documentation searched (classification system followed by classification symbols)

IPC 7 H04N

Documentation searched other than minimum documentation to the extent that such documents are included in the fields searched

Electronic data base consulted during the international search (name of data base and, where practical, search terms used)

EPO-Internal, INSPEC

## C. DOCUMENTS CONSIDERED TO BE RELEVANT

Category *	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
X	US 4 723 161 A (KOGA TOSHIO) 2 February 1988 (1988-02-02) column 5, line 25 - line 40 figures 3-6	1-35
X	EP 0 595 403 A (PHILIPS ELECTRONIQUE LAB ;KONINKL PHILIPS ELECTRONICS NV (NL)) 4 May 1994 (1994-05-04) column 2, line 1 - line 23 column 3, line 3 -column 6, line 22 figures 1,2 -/-	1-35

☒ Further documents are listed in the continuation of box C.☒ Patent family members are listed in annex.

## \* Special categories of cited documents:

\*A\* document defining the general state of the art which is not considered to be of particular relevance

\*E\* earlier document but published on or after the international filing date

\*L\* document which may throw doubts on priority claim(s) or which is cited to establish the publication date of another citation or other special reason (as specified)

\*O\* document referring to an oral disclosure, use, exhibition or other means

\*P\* document published prior to the international filing date but later than the priority date claimed

\*T\* later document published after the international filing date or priority date and not in conflict with the application but cited to understand the principle or theory underlying the invention

\*X\* document of particular relevance; the claimed invention cannot be considered novel or cannot be considered to involve an inventive step when the document is taken alone

\*Y\* document of particular relevance; the claimed invention cannot be considered to involve an inventive step when the document is combined with one or more other such documents, such combination being obvious to a person skilled in the art.

\*Z\* document member of the same patent family

Date of the actual completion of the international search

2 August 2002

Date of mailing of the international search report

16/08/2002

Name and mailing address of the ISA

European Patent Office, P.B. 5818 Patentlaan 2  
NL - 2280 HV Rijswijk  
Tel. (+31-70) 340-2040, Tx. 31 651 epo nl,  
Fax: (+31-70) 340-3016

Authorized officer

Lombardi, G

## INTERNATIONAL SEARCH REPORT

International Application No

PCT/EP 01/11393

## C.(Continuation) DOCUMENTS CONSIDERED TO BE RELEVANT

Category *	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
X	JOHNSON A W ET AL: "FREQUENCY SCALABLE VIDEO CODING USING THE MDCT" PROCEEDINGS OF THE INTERNATIONAL CONFERENCE ON ACOUSTICS, SPEECH AND SIGNAL PROCESSING (ICASSP). I. IMAGE AND MULTIDIMENSIONAL SIGNAL PROCESSING. ADELAIDE, APR. 19 - 22, 1994, NEW YORK, IEEE, US, vol. 5 CONF. 19, 19 April 1994 (1994-04-19), pages V-477-V-480, XP000533761 ISBN: 0-7803-1776-9 figure 4 ---	1-35
A	WU F ET AL: "DCT-prediction based progressive fine granularity scalable coding" PROCEEDINGS. INTERNATIONAL CONFERENCE ON IMAGE PROCESSING, 10 September 2000 (2000-09-10), pages 556-559, XP002165186 page 557, paragraph 2 page 557 -page 558, paragraph 3 figures 1,2 -----	1-35

## INTERNATIONAL SEARCH REPORT

Information on patent family members

International Application No

PCT/EP 01/11393

Patent document cited in search report		Publication date	Patent family member(s)	Publication date
US 4723161	A	02-02-1988	JP 2103366 C	22-10-1996
			JP 7118807 B	18-12-1995
			JP 61214883 A	24-09-1986
			JP 2103367 C	22-10-1996
			JP 7118808 B	18-12-1995
			JP 61214885 A	24-09-1986
			CA 1251276 A1	14-03-1989
			GB 2173067 A , B	01-10-1986
EP 0595403	A	04-05-1994	FR 2697393 A1	29-04-1994
			AT 176842 T	15-03-1999
			AU 674884 B2	16-01-1997
			AU 5030893 A	12-05-1994
			CA 2109138 A1	29-04-1994
			CN 1088041 A , B	15-06-1994
			DE 69323523 D1	25-03-1999
			DE 69323523 T2	02-09-1999
			EP 0595403 A1	04-05-1994
			ES 2130214 T3	01-07-1999
			FI 934714 A	29-04-1994
			JP 6233137 A	19-08-1994
			US 5500677 A	19-03-1996